

TSHWANE UNIVERSITY OF TECHNOLOGY
&
UNIVERSITÉ DE VERSAILLES SAINT-QUENTIN
&
MENSIA TECHNOLOGIES

Research Report: Using Riemannian geometry for SSVEP-based Brain Computer Interface

Authors:

Emmanuel KALUNGA

Sylvain CHEVALLIER

Quentin BARTHÉLEMY

Affiliations:

Department of Electrical
Engineering, F'SATIE, TUT,
Pretoria, South Africa

Laboratoire d'Ingénierie des
Systèmes de Versailles, UVSQ,
Velizy, France

Mensia Technologies, ICM,
Hôpital de la Pitié-Salpêtrière,
Paris, France

February 25, 2015



Using Riemannian geometry for SSVEP-based Brain Computer Interface

Emmanuel K. Kalunga* Sylvain Chevallier[†]

Quentin Barthélemy[‡]

February 25, 2015

Abstract

Riemannian geometry has been applied to Brain Computer Interface (BCI) for brain signals classification yielding promising results. Studying electroencephalographic (EEG) signals from their associated covariance matrices allows a mitigation of common sources of variability (electronic, electrical, biological) by constructing a representation which is invariant to these perturbations. While working in Euclidean space with covariance matrices is known to be error-prone, one might take advantage of algorithmic advances in information geometry and matrix manifold to implement methods for Symmetric Positive-Definite (SPD) matrices. This paper proposes a comprehensive review of the actual tools of information geometry and how they could be applied on covariance matrices of EEG. In practice, covariance matrices should be estimated, thus a thorough study of all estimators is conducted on real EEG dataset. As a main contribution, this paper proposes an online implementation of a classifier in the Riemannian space and its subsequent assessment in Steady-State Visually Evoked Potential (SSVEP) experimentations.

Keywords:

Riemannian geometry, Online, Asynchronous, Brain-Computer Interfaces, Steady State Visually Evoked Potentials.

*Department of Electrical Engineering, French South African Institute of Technology, Tshwane University of Technology, Pretoria 0001, South Africa. emmanuelkalunga.k@gmail.com

[†]Laboratoire d'Ingénierie des Systèmes de Versailles, Université de Versailles Saint-Quentin, Velizy 78140, France. sylvain.chevallier@uvsq.fr

[‡]Mensia Technologies, ICM, Hôpital de la Pitié-Salpêtrière, 75013 Paris, France. quentin.barthelemy@mensiatech.com

1 Introduction

Human-machine interactions without relying on muscular capabilities is possible with Brain-Computer Interfaces (BCI) [1, 2, 3]. They are the focus of a large scientific interest [4, 5, 6, 7, 8, 9], especially those based on electroencephalography (EEG) [10, 11]. From the large literature based on the BCI competition datasets [12, 13, 14], one can identify the two most challenging BCI problems: on one hand, the inter-individual variability plagues the models and lead to BCI-inefficiency effect [15, 16, 17], on the other hand the intra-individual changes calls for the development of online algorithms and adaptive systems following the evolution of the subject’s brain waves [18, 19, 20, 21, 22]. To alleviate these variations, several signal processing and machine learning techniques have been proposed, such as filtering, regularization or clustering [23, 24, 25] without marking out an obvious “best candidate” methodology.

A common vision is shared by all the most efficient approaches to reduce signal variabilities: they are applied on covariance matrices instead of working in the input signal space. Common Spatial Pattern (CSP) [26, 27, 28, 29], which is the most known preprocessing technique in BCI, try to maximize the covariance of one class while minimizing the covariance of the other one. Similarly, Principal Components Analysis (PCA) [12, 13], also applied for spatial filtering in BCI, is based on the estimation of covariance matrices. Canonical Correlation Analysis (CCA) [30] is another example of a technique relying on covariance estimates successfully applied on EEG for spatial filtering [31, 21, 24]. Given two sets of signals, CCA aims at finding the projection space that maximizes their cross-covariance while jointly minimizing their covariance. Covariance matrices are also found in classifiers such as the Linear Discriminant Analysis (LDA), which is largely used in BCI. In all cases, they are handled as elements of an Euclidean space. However, being Symmetric and Positive-Definite (SPD), covariance matrices lie on a subset of the Euclidean space, with reduced dimensionality and specific properties, the *Riemannian manifold*. Considering covariance matrices in their original space would reduce the search area for an optimization problem [32]. As Riemannian manifolds inherently define a metric, the distance between SPD matrices takes into account the space where they lie on; approximating it to an Euclidean space introduce inaccuracies and result in ill-conditioned matrices.

Recently, studies have been done to consider covariance matrices obtained from multichannel brain signals in their original space [33, 34]. Covariance matrices are the input features of the BCI system and the classifier algorithms rely on Riemannian metric for partitioning the feature space. The authors propose to build specific covariance matrices in order to emphasize the spatial and temporal information of the multichannel brain signals. The outcome of this approach is a simple processing toolchain, which achieve

state-of-the-art classification performances.

This paper introduces an online version of the *minimum distance to Riemannian mean* algorithm [34], with an application on Steady-State Visually Evoked Potential (SSVEP) recordings. In SSVEP, the subjects concentrate on stimuli blinking with fixed frequencies; depending on the focus of their attention, brain waves will arise with the same phase and frequency than the stimulus chosen by the subject. The proposed online implementation improves the performances in the BCI task, as the parameters of the classifier adapt to new, unlabeled samples. Such adaptation takes into account the qualitative changes in the brain responses and the acquired signals, caused respectively by internal and external factors, e.g. user fatigue and slight variations in experimental settings. The proposed online implementation is similar to the unsupervised or semi-supervised learning scheme proposed in [35, 36]; that has the potential of shortening (or even removing) the calibration phase. The proposed algorithm applies a similar approach to dynamic stopping criterion used in [37] to increase the speed of the BCI system. This approach allows to dynamically determine the trial length and ensure robustness in classification results. When working with covariance matrices, a crucial point is to correctly estimate the covariance when the number of samples is small or heavily corrupted by noise. Several approaches have been proposed to build the covariance matrices, relying on normalization or regularization of the sample covariances. To assess the quality of the covariance matrices obtained from EEG samples, a comparative study of these estimators is conducted.

Hence, the contributions of this work are:

- a comprehensive review of the literature on Riemannian geometry applied to EEG and time-series,
- an online and asynchronous classification algorithm for SSVEP-based BCI,
- a thorough analysis of the covariance estimators and their impact on tools derived from information geometry.

The paper is divided as follows: Section 2 reviews application of Riemannian geometry to machine learning. Section 3 presents concepts of Riemannian geometry relevant to this work and estimators of covariance. In Section 4, the proposed classification algorithm for online SSVEP is introduced and the experimental results are presented in Section 5.

2 State of the Art

Information geometry provides useful tools for various machine learning and optimization problems. In machine learning, SPD matrices have been used

in various applications where features and data are only considered in the Euclidean space. Indeed, covariance matrices lie in the space of SPD matrices which is a subset of the Euclidean space when considered with the scalar product. But the same space of SPD matrices, endowed with a differential structure, induces a Riemannian manifold.

Riemannian geometry can improve machine learning algorithms, taking explicitly into consideration the underlying structure of the considered space. Three kind of approaches in the literature uses the data geometry in machine learning. The first one relies on the mapping of the Riemannian manifold onto an Euclidean vector space. One such mapping, called logarithmic mapping, exist between the manifold and its tangent space, which is an Euclidean space, and has been used in classification task for BCI [38, 39]. Other kernels have been applied successfully to this end: Stein kernel, Log-Euclidean kernels as well as their normalized versions [40]. The main idea is to map the input data to a high dimensional feature space, providing a rich and hopefully linearly separable representation. The so-called kernel trick is to provide a kernel function, which computes an inner product in the feature space directly from points lying in the input space, defining a Reproducing Kernel Hilbert Space (RKHS). The family of kernels defined on the Riemannian manifold allows implementing extension of all kernel-based methods, such as SVM, kernel-PCA or kernel k -means [41]. Apart from the kernel approaches, once the data are mapped onto a vector space, any machine learning algorithm working in Euclidean space, such as LDA, could be applied [34].

A second kind of machine learning approach exploit the underlying geometry of the data. Instead of mapping the data to an Euclidean space, either a tangent space or a RKHS, the algorithms are adapted to Riemannian space. For instance, sparse coding algorithm have been adapted to Riemannian manifold, using the geodesic distance to estimate the data point and its sparse estimate [42]. Similarly nonlinear dimensionality reduction techniques have been adapted to Riemannian manifold, such as Laplacian Eigenmaps (LE), Locally Linear Embedding (LLE), and Hessian LLE. This adaptation was used to cluster data using their pdfs [43] or covariance matrices [44] as features. Another example is the adaptation of interpolation and filtering of data to Riemannian space performed in [45], where an affine-invariant Riemannian metric is also proposed to offer a geodesically complete manifold i.e a manifold with no edge and no singular point that can be reached in finite time.

In the last kind of approach, instead of adapting existing algorithm from Euclidean to Riemannian geometry, new algorithms are developed directly for Riemannian manifolds. The *minimum distance to Riemannian mean* (MDRM) relies on a Riemannian metric to implement a multi-class classifier and have been applied on EEG. New EEG trials are assigned to the class whose average covariance matrix is the closest to the trial covariance

matrix [34]. The MDRM classification can be preceded by a filtering of covariance matrices, like in [33] where covariance matrices are filtered with LDA component in the tangent space, then brought back to the Riemannian space for classification with MDRM. Another example is the *Riemannian Potato* [46], an unsupervised and adaptive artifact detection method, providing an online adaptive EEG filtering (i.e outliers removal). Incoming signals are rejected if their covariance matrix lies beyond a predefined distance z -score from the mean covariance matrix, computed from a sliding window. With the same objective of achieving robustness to noise that affect covariance matrices, Riemannian geometry is used to solve divergence functions of pdfs [47]. This allows to reformulate the CSP as the maximization of the divergence between the distributions of data from two different classes corresponding to two cognitive states [48, 49]. Using the *beta divergence* the obtained CSP is robust to outliers in sample covariance matrices and this algorithm is successfully applied to EEG filtering for BCI. Riemannian metrics are also used for EEG channel selection [50] and the selection of the most discriminatory spatial filters in CSP [51].

Applications of Riemannian geometry to BCI mentioned thus far are focusing on motor imagery (MI) paradigm. In MI experiment, the subject is asked to imagine a movement (usually hand, feet or tongue), generating Event-Related Synchronization and Desynchronization (ERD/ERS) in pre-motor brain area. Riemannian BCI is well suited for MI experiment as the spatial information linked with synchronization is directly embedded in covariance matrices obtained from multichannel recordings. However, for BCI that rely on Evoked Potential such as SSVEP or Event Related Potential (ERP), as P300, both frequential and temporal information are needed; the spatial covariance matrix does not contain these information. To apply Riemannian geometry to SSVEP and ERP, the sample covariance matrices can be defined from a rearrangement of the recorded data. The rearrangement is done such that the temporal or frequency information are captured [52]. With similar motivations, [53] and [54] defined a new Riemannian distance between SPD matrices that would take into account a weighting factor on matrices. They use this new distance as a dissimilarity between weighted matrices of power spectral density to classify EEG into different sleep state by k -nearest neighbors.

3 Covariance Matrices and their Geometry

This section presents some formal definitions for the information geometry concepts used in this paper. The link with the covariance matrices is explicated in Section 3.2, along with the covariance estimators proposed in the literature.

3.1 Riemannian Manifold

An m -dimensional *manifold* \mathcal{M} is a Hausdorff space for which every point has a neighborhood that is homeomorphic to an open subset of \mathbb{R}^m [55, 56]. When a tangent space is defined at each point, \mathcal{M} is called a differential manifold. A *geodesic* γ is the shortest smooth curve between two points, Σ_1 and Σ_2 . The tangent space $T_\Sigma \mathcal{M}$ at point Σ is the vector space spanned by the tangent vectors of all geodesics on \mathcal{M} passing through Σ . A *Riemannian* manifold is a manifold endowed with an inner product defined on the tangent space, which varies smoothly from point to point.

For the rest of this paper, we will restrict to the analysis of the manifold \mathcal{M}_C of the $C \times C$ symmetric positive definite matrices, defined as:

$$\mathcal{M}_C = \{\Sigma \in \mathbb{R}^{C \times C} : \Sigma = \Sigma^\top \text{ and } x^\top \Sigma x > 0, \forall x \in \mathbb{R}^C \setminus \{0\}\} .$$

The tangent space $T_\Sigma \mathcal{M}_C$ is identified to the Euclidean space of symmetric matrices:

$$\mathcal{S}_C = \{\Theta \in \mathbb{R}^{C \times C} : \Theta = \Theta^\top\} .$$

The dimension of the manifold \mathcal{M}_C , and its tangent space $T_\Sigma \mathcal{M}_C$, is $m = C(C+1)/2$.

The mapping from a point Θ_i of the tangent space to the manifold is called the exponential mapping $\text{Exp}_\Sigma(\Theta_i): T_\Sigma \mathcal{M}_C \rightarrow \mathcal{M}_C$ and is defined as:

$$\text{Exp}_\Sigma(\Theta_i) = \Sigma^{\frac{1}{2}} \text{Exp}(\Sigma^{-\frac{1}{2}} \Theta_i \Sigma^{-\frac{1}{2}}) \Sigma^{\frac{1}{2}} . \quad (1)$$

Its inverse mapping, from the manifold to the tangent space is the logarithmic mapping $\text{Log}_\Sigma(\Sigma_i): \mathcal{M}_C \rightarrow T_\Sigma \mathcal{M}_C$ and is defined as:

$$\text{Log}_\Sigma(\Sigma_i) = \Sigma^{\frac{1}{2}} \text{Log}(\Sigma^{-\frac{1}{2}} \Sigma_i \Sigma^{-\frac{1}{2}}) \Sigma^{\frac{1}{2}} . \quad (2)$$

$\text{Exp}(\cdot)$ and $\text{Log}(\cdot)$ are the the matrix exponential and logarithm respectively. The computation of these operators is straightforward for SPD matrices of \mathcal{M}_C . They are obtained from their eigenvalue decomposition (EVD):

$$\begin{aligned} \Sigma &= U \text{diag}(\lambda_1, \dots, \lambda_C) U^\top , \\ \text{Exp}(\Sigma) &= U \text{diag}(\log(\lambda_1), \dots, \log(\lambda_C)) U^\top , \\ \text{Log}(\Sigma) &= U \text{diag}(\exp(\lambda_1), \dots, \exp(\lambda_C)) U^\top , \end{aligned}$$

where $\lambda_1, \dots, \lambda_C$ are the eigenvalues and U the matrix of eigenvectors of Σ . As any SPD matrix can be diagonalized with strictly positive eigenvalues, $\text{Log}(\cdot)$ is always defined. Similarly the square root $\Sigma^{\frac{1}{2}}$ is obtained as:

$$\Sigma^{\frac{1}{2}} = U \text{diag}(\lambda_1^{\frac{1}{2}}, \dots, \lambda_C^{\frac{1}{2}}) U^\top ,$$

and is unique. The same goes for $\Sigma^{-\frac{1}{2}}$.

The tangent vector of the geodesic $\gamma(t)$ between Σ_1 and Σ_2 , where $\gamma(0) = \Sigma_1$ and $\gamma(1) = \Sigma_2$ is defined as $v = \overrightarrow{\Sigma_1 \Sigma_2} = \text{Log}_{\Sigma_1}(\Sigma_2)$. A Riemannian distance between Σ_1 and Σ_2 can thus be defined as [57]:

$$\delta(\Sigma_1, \Sigma_2) = \|\text{Log}(\Sigma_1^{-1} \Sigma_2)\|_F = \left[\sum_{c=1}^C \log^2 \lambda_c \right]^{1/2}, \quad (3)$$

where λ_c , $c = 1, \dots, C$, are the eigenvalues of $\Sigma_1^{-1} \Sigma_2$. From Eq. (3), the geometric mean of I points Σ_i on the manifold, $i = 1, \dots, I$, can be defined as the point that minimizes the sum of squared distances to all Σ_i :

$$\mu(\Sigma_1, \dots, \Sigma_I) = \arg \min_{\Sigma \in \mathcal{M}_C} \sum_{i=1}^I \delta^2(\Sigma_i, \Sigma). \quad (4)$$

This geometric mean has no closed form, and can be computed iteratively [58].

3.2 Covariance Matrix Estimation

Let $x_n \in \mathbb{R}^C$, $n = 1, \dots, N$, denotes a sample of a multichannel EEG trial recorded on C electrodes. N is the trial length. Let $X \in \mathbb{R}^{C \times N}$ be the EEG trial such as $X = [x_1, \dots, x_N]$. Under the hypothesis that all N samples x_n are randomly drawn from a distribution, it follows that X is a variable of random vectors and its expected vector is $\omega = E\{X\}$ [59]. The covariance matrix of the random vector X is defined by $\Sigma = E\{(X - \omega)(X - \omega)^\top\}$ and is unknown, thus an estimate $\hat{\Sigma}$ should be computed. The choice of the appropriate estimator is crucial to verify that the obtained covariance matrices fulfill the following properties: they should be accurate, SPD and well-conditioned. The last property requires that the ratio between the maximum and minimum singular value is not too large. Moreover, to ensure the computational stability of the algorithm, the estimator should provide full-rank matrices, and its inversion should not amplify estimation errors.

3.2.1 Sample Covariance Matrix Estimator

The most usual estimator is the empirical *sample covariance matrix* (SCM), defined as:

$$\begin{aligned} \hat{\Sigma}_{\text{scm}} &= \frac{1}{N-1} \sum_{n=1}^N (x_n - \bar{x})(x_n - \bar{x})^\top \\ &= \frac{1}{N-1} X \left(\mathbb{I}_N - \frac{1}{N} \mathbb{1}_N \mathbb{1}_N^\top \right) X^\top, \end{aligned} \quad (5)$$

where \bar{x} is the sample mean vector $\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n$. In the matrix notation, \mathbb{I}_N is the $N \times N$ identity matrix and $\mathbb{1}_N$ is the vector $[1, \dots, 1]$. The SCM

is often normalized as:

$$\hat{\Sigma}_{\text{nscm}} = \frac{C}{N} \sum_{n=1}^N \frac{(x_n - \bar{x})(x_n - \bar{x})^\top}{\sigma_{x_n}^2}, \quad (6)$$

$\sigma_{x_n}^2$ is the inter-channel variance at time n . Other normalization techniques could be used.

This estimation is fast and computationally simple. However when $C \approx N$, the SCM is not a good estimator of the true covariance. In the case $C > N$, the SCM is not even full rank.

3.2.2 Shrinkage Covariance Matrix Estimators

To overcome these shortcomings, the shrinkage estimators have been developed as a weighted combination of the SCM and a target covariance matrix, which is often chosen to be close to the identity matrix, *i.e.* resulting from almost independent variables of unit variance.

$$\hat{\Sigma}_{\text{shrink}} = \kappa \Gamma + (1 - \kappa) \hat{\Sigma}_{\text{scm}}, \quad (7)$$

where $0 \leq \kappa < 1$. This estimator provides a regularized covariance that outperforms the empirical $\hat{\Sigma}_{\text{scm}}$ for small sample size, that is $C \approx N$. The shrinkage estimator has the same eigenvectors as the SCM, but the extreme eigenvalues are modified *i.e.* the estimator is shrunk or elongated toward the average.

The different shrinkage estimators differ in their definition of the target covariance matrix Γ . Ledoit and Wolf [60] have proposed $\Gamma = v \mathbb{I}_C$, with $v = \text{Tr}(\hat{\Sigma}_{\text{scm}} \mathbb{I}_C)$. Blankertz [61] defines Γ also as $v \mathbb{I}_C$ but with $v = \frac{\text{Tr}(\hat{\Sigma}_{\text{scm}})}{m}$, m being the dimension of the covariance matrices space. Schäfer proposes several ways of defining Γ depending on the observed $\hat{\Sigma}_{\text{scm}}$ [62].

3.2.3 Fixed Point Covariance Matrix Estimator

The Fixed Point Covariance Matrix [63] is based on the maximum likelihood estimator $\hat{\ell}$ which is a solution to the following equation:

$$\hat{\Sigma}_{\text{fp}} = \hat{\ell} = \frac{C}{N} \sum_{n=1}^N \left(\frac{(x_n - \bar{x})(x_n - \bar{x})^\top}{(x_n - \bar{x})^\top \hat{\ell}^{-1} (x_n - \bar{x})} \right). \quad (8)$$

As there is no closed form expression to Eq. (8), it can be written as a function of $\hat{\ell}$: $g(\hat{\ell}) = \hat{\Sigma}_{\text{fp}}$. g admits a single *fixed point* $\hat{\ell}^*$, where $g(\hat{\ell}^*) = \hat{\ell}^*$, which is a solution to Eq. (8). Using $\hat{\ell}_0 := \hat{\Sigma}_{\text{nscm}}$ as the initial value of $\hat{\ell}$, it is solved recursively as $\hat{\ell}_t \xrightarrow[t \rightarrow \infty]{} \hat{\ell}^*$.

4 Online Adaptation of the Riemannian Classifier

Concerning Riemannian classification of SSVEP, the offline methodology has been explained in [52]. In this paper, we propose an online classifier for SSVEP, composed of an offline training step and an online and asynchronous test step. This analysis is performed for each subject independently.

4.1 Offline Riemannian Classification

The proposed classifier relies on the Minimum Distance to Riemannian Mean (MDRM) introduced in [33] and extended in [52] for possible applications on SSVEP signals. Let consider an experimental SSVEP setup with F stimulus blinking at F different frequencies. It is a multiclass classification with $K = F + 1$ classes: one class per stimulus and one resting state class. The covariance matrices are estimated from a modified version of the input signal X :

$$X \in \mathbb{R}^{C \times N} \rightarrow \begin{bmatrix} X_{\text{freq}_1} \\ \vdots \\ X_{\text{freq}_F} \end{bmatrix} \in \mathbb{R}^{FC \times N}, \quad (9)$$

where X_{freq_f} is the input signal X band-pass filtered around frequency freq_f , $f = 1, \dots, F$. Thus the resulting covariance matrix $\hat{\Sigma}$ belongs to $\mathbb{R}^{FC \times FC}$. Henceforth, all EEG signals will be considered as filtered and modified by Eq. (9).

From I labelled training trials $\{X_i\}_{i=1}^I$ recorded per subject, K centers of classes $\Sigma_\mu^{(k)}$ are estimated using Algorithm 1. When an unlabelled test trial Y is given, it is classified as belonging to the class whose center $\Sigma_\mu^{(k)}$ is the closest to the trial's covariance matrix (Algorithm 2, step 2).

Algorithm 1 Offline Estimation of Riemannian Centers of Classes

Inputs: $X_i \in \mathbb{R}^{FC \times N}$, for $i = 1, \dots, I$, a set of labelled trials.

Inputs: $\mathcal{I}(k)$, a set of indices of trials belonging to class k .

Output: $\Sigma_\mu^{(k)}$, $k = 1, \dots, K$, centers of classes.

- 1: Compute covariance matrices $\hat{\Sigma}_i$ of X_i
 - 2: **for** $k = 1$ **to** K **do**
 - 3: $\Sigma_\mu^{(k)} = \mu(\hat{\Sigma}_i : i \in \mathcal{I}(k))$, Eq. (4)
 - 4: **end**
 - 5: **return** $\Sigma_\mu^{(k)}$
-

Algorithm 2 Minimum Distance to Riemannian Mean

Inputs: $\Sigma_\mu^{(k)}$, K centers of classes from Algorithm 1.

Input: $Y \in \mathbb{R}^{FC \times N}$, an unlabelled test trial.

Output: k^* , the predicted label of Y .

- 1: Compute covariance matrix $\hat{\Sigma}$ of Y
 - 2: $k^* = \operatorname{argmin}_k \delta(\hat{\Sigma}, \Sigma_\mu^{(k)})$
 - 3: **return** k^*
-

Algorithm 3 Curve-based Online Classification

Inputs: hyper-parameters w , Δn , D , and ϑ .

Inputs: $\Sigma_\mu^{(k)}$, $k = 1, \dots, K$, centers of classes from Algorithm 1 (offline training).

Inputs: Online EEG recording $\chi(n)$.

Output: $\tilde{k}(n)$, online predicted class.

- 1: $d = 1$
 - 2: **for** $n = w$ **to** \mathcal{N} **step** Δn
 - 3: Epoch X_d , Eq. (10), and classify it with Algorithm 2
 - 4: **if** $d \geq D$
 - 5: Find the most recurrent class in $\mathcal{K} = k_{j \in \mathcal{J}(d)}^*$: $\bar{k} = \arg \max_k \rho(k)$, Eq. (11)
 - 6: **if** $\rho(\bar{k}) > \vartheta$
 - 7: Compute $\tilde{\delta}_{\bar{k}}$, Eq. (12)
 - 8: **if** $\tilde{\delta}_{\bar{k}} < 0$
 - 9: **return** $\tilde{k} = \bar{k}$
 - 10: **end**
 - 11: **end**
 - 12: **end**
 - 13: $d = d + 1$
 - 14: **end**
-

4.2 Curve-based Online Classification

In offline synchronous BCI paradigm, cue onset are used as reference for the localization of a brain response, e.g. for an evoked potential. Nonetheless most of the BCI applications are online and asynchronous; cue onsets are not known, thus designing online version of BCI algorithms, even well documented ones, is not a trivial task. The approach introduced here identifies a period (*i.e.* interval) in the online EEG $\chi \in \mathbb{R}^{FC \times \mathcal{N}}$, where \mathcal{N} is the number of recorded samples, associated with a high probability (above threshold) of observing an SSVEP at a specific frequency, as illustrated in Algorithm 3.

To locate this interval, we focus on the last D recorded EEG overlapping epochs $\{X_i \in \mathbb{R}^{FC \times w}\}_{j \in \mathcal{J}(d)}$, with the set of indices $\mathcal{J}(d) = d - D + 1, \dots, d - 1, d$; where d is the index of the current epoch X_d in the online recording $\chi(n)$. Epochs have size w , and the interval between two consecutive epochs is Δn , with $w > \Delta n$:

$$X_d = \chi(n - w, \dots, n) . \quad (10)$$

To obtain the first D epochs $X_j \in \mathcal{J}(d)$, at least $w + (D - 1) \Delta n$ samples of χ should be recorded (step 4).

The classification outputs $k_{j \in \mathcal{J}(d)}^*$ obtained by applying Algorithm 2 (step 3) on $X_j \in \mathcal{J}(d)$ are stored in a vector \mathcal{K} . The class that occurs the most in \mathcal{K} (step 5), with an occurrence probability $\rho(k)$ above a defined threshold ϑ , is considered to be the class, denoted \bar{k} , of the ongoing EEG recording $\chi(n)$. Vector ρ is defined as:

$$\rho(k) = \frac{\#\{k_{j \in \mathcal{J}(d)}^* = k\}}{D} , \text{ for } k = 1, \dots, K, \quad (11)$$

and $\bar{k} = \arg \max_k \rho(k) .$

If ϑ is not reached within the last D epochs, the classification output is held back, and the sliding process continues until ϑ is reached. In the last D epochs, once a class \bar{k} has been identified, a curve direction criterion is introduced to enforce the robustness of the result. For class \bar{k} to be validated, this criterion requires that the direction taken by the displacement of covariance matrices $\hat{\Sigma}_{j \in \mathcal{J}(d)}$ be toward the center of class $\Sigma_{\mu}^{(\bar{k})}$. Hence $\tilde{\delta}_{\bar{k}}$, the sum of gradients (*i.e.* differentials) of the curve made by distances from $\hat{\Sigma}_{j \in \mathcal{J}(d)}$ to $\Sigma_{\mu}^{(\bar{k})}$ should be negative (step 8):

$$\tilde{\delta}_{\bar{k}} = \sum_{j \in \mathcal{J}(d)} \frac{\Delta \delta_{\bar{k}}(j)}{\Delta j} = \sum_{j=d-D+2}^d \delta_{\bar{k}}(j) - \delta_{\bar{k}}(j-1) < 0 \quad (12)$$

with $\delta_{\bar{k}}(j) = \frac{\delta(\hat{\Sigma}_j, \Sigma_{\mu}^{(\bar{k})})}{\sum_{k=1}^K \delta(\hat{\Sigma}_j, \Sigma_{\mu}^{(k)})} .$

The occurrence criterion is inspired by the dynamic stopping of [37]; there is no fixed trial length for classification. The occurrence criterion ensures that the detected user intention is unaffected by any short time disturbances due to noise or subject’s inattention, as presented in Algorithm 3. This approach offers a good compromise to obtain robust results within a short and flexible time.

5 Experimental Validation

Covariance matrix estimators, Algorithms 2 and 3 are applied to SSVEP signals for offline and online analysis. This section presents the analysis and results obtained.

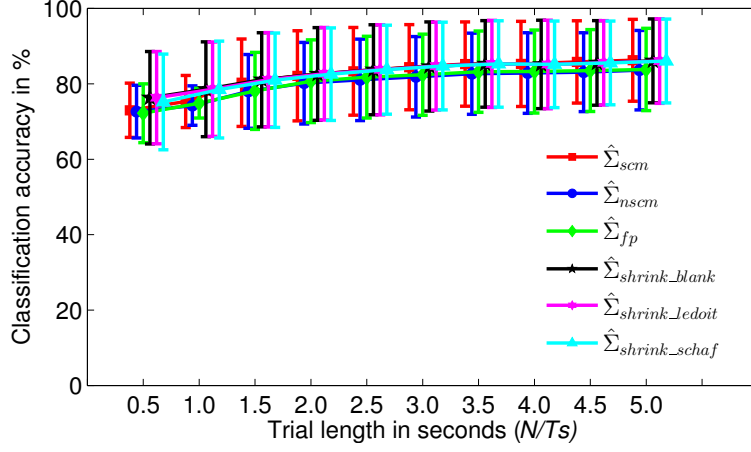
5.1 Data Description

The signals are recorded from 12 subjects during an SSVEP experiment. EEG are measured on $C = 8$ channels: O_Z , O_1 , O_2 , PO_Z , PO_3 , PO_4 , PO_7 , and PO_8 . The ground and the reference electrodes were placed respectively on F_Z and the right hear mastoid. The acquisition rate is $T_s = 256$ Hz on a gTec MobiLab amp. The subjects are presented with $F = 3$ visual target stimuli blinking respectively at $\text{freq} = 13\text{Hz}$, 17Hz and 21Hz . It is a $K = 4$ classes BCI setup made of the $F = 3$ stimulus classes and one resting class (no-SSVEP). In a session, which lasts 5 minutes, 32 trials are recorded: 8 for each visual stimulus and 8 for the resting class. The number of sessions recorded per subject varies from 2 to 5. Thus the longest EEG recorded for a single subject is 25 minutes or 160 trials. The trial length is 6 seconds, that is $N = 6 \times T_s = 1536$ samples. Since data are rearranged as detailed in (9), trials $X \in \mathbb{R}^{FC \times N}$, where $FC = 24$ corresponding to 8 channels times 3 stimulus frequencies. For each subject, a test set is made of 32 trials while the remaining trials (which might vary from 32 to 128) make up for the training set.

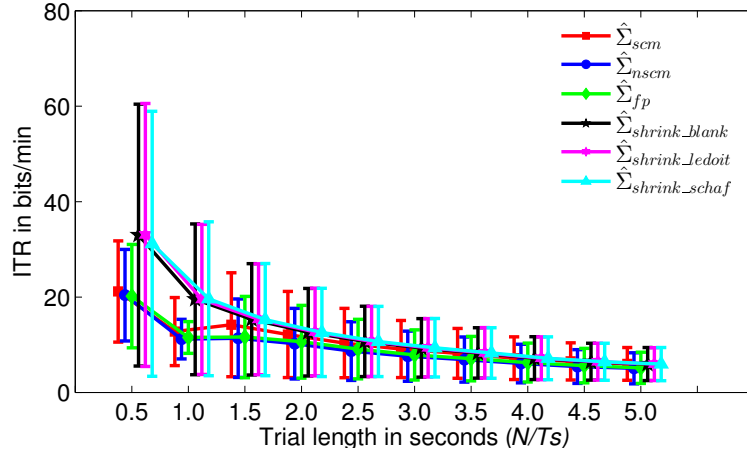
5.2 Covariance Estimators Comparison

In this section, the effectiveness of covariance matrix estimators is evaluated for SSVEP signals. The evaluation is done in terms of classification accuracy, information transfer rate (ITR), and integrated discrimination improvement (IDI), obtained by each estimator (see Section 3.2) while using the offline MDRM classifier. The different conditioning of covariance matrices are also investigated.

A bootstrapping with 1000 replications is performed to assess the performances of each estimator. Estimators are compared on 10 trial lengths $t \in \{0.5, 1.0, \dots, 5.0\}$ seconds, as these are known to affect the estimators performance. Here $N \in \{128, 256, \dots, 1280\}$ is computed as $N = t \times T_s$.



(a)



(b)

Figure 1: Comparison of covariance estimators in terms of classification accuracy and Information Transfer Rate (ITR) obtained with MDRM with increasing EEG trial length. For each trial length, the average accuracy and ITR across all subjects and across all replications is shown. Bars indicate the standard deviation. (a) Accuracy in percentage, (b) ITR in bits/min.

Figure 1a shows the classification accuracies of each estimator. The increase in the accuracy can be attributed to the fact that the relevant patterns in EEG accumulate with the trial length, producing better estimation of the covariance matrices. This is known to be particularly true for the SCM estimator and it could be seen in Figure 1a. It appears that shrinkage estimators (especially Ledoit and Schäfer) are less affected by the reduction of epoch sizes than the other estimators. This is a direct consequence of the regularization between the sample covariance matrices and the targeted

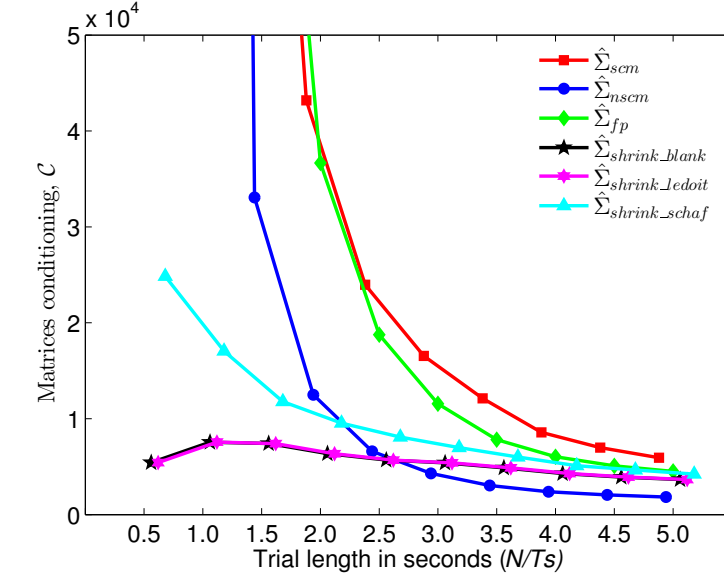
(expected) covariance matrix of independent variables. The information of Figure 1a is congruent with Figure 1b; ITR increases as the trial length decreases. The transfer rate is higher for shrinkage estimators with shorter epoch sizes, while maintaining correct accuracy.

For computational purposes, it is important to look at the matrix conditioning. Figure 2a shows the ratio \mathcal{C} between the largest and smallest eigenvalues: in well-conditioned matrices, \mathcal{C} is small. Shrinkage estimators offer better conditioned matrices while the SCM, NSCM, and Fixed Point matrices are ill-conditioned below 2 seconds of trial length, and may result in singular matrices.

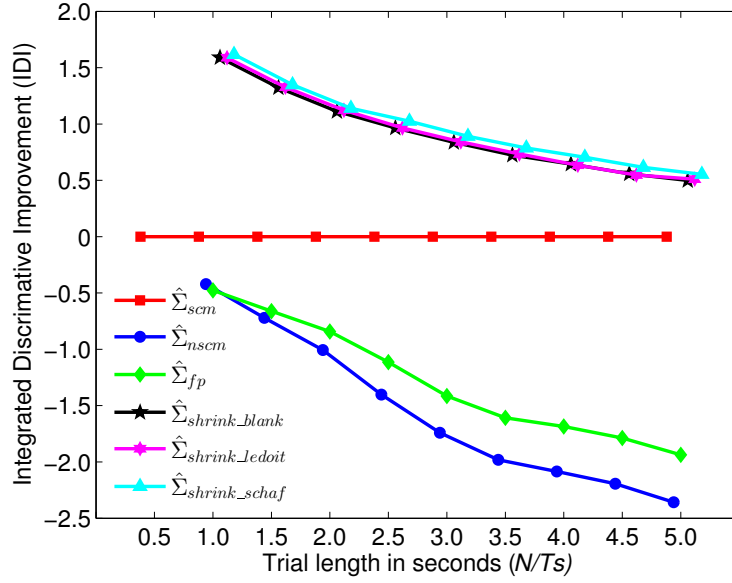
On Figure 2b, the Integrated Discrimination Improvement (IDI), as defined in [64], is computed for the different estimators and trial lengths. The SCM is used as a reference for improvement, as this is the most popular estimator in the literature. Negative IDI means a deterioration in the method discrimination ability. It is clear that shrinkage estimators increase the discrimination power of the classifier. However, despite being more complex than the SCM, the NSCM and the Fixed Point estimators decrease the discrimination ability of classifiers. From Figures 1a and 2b, it is apparent that the difference in performance between the SCM and shrinkage estimators reduces as the trial length increases. The simplicity of the SCM plays a favorable role: it is an attractive method for longer trials. However for trial lengths compatible with a reasonable ITR for man-machine interaction (< 5 sec), the shrinkage estimators offer better performances, with the Schafer estimator being the best. The p -values under the hypothesis that there is no improvement (*i.e.* IDI = 0) from one estimator to another are all inferior to 10^{-47} , hence the improvement is significant. It should be noted that the estimation of covariance matrices is a trade-off between the quality of the estimate and the computation time required; this should be considered for real-time processing.

5.3 Effect of Outliers on Centers Estimation

Outliers can affect the offline training of the K centers of class $\Sigma_{\mu}^{(k)}$ by Algorithm 1. Figure 3 shows representations of training covariance matrices Σ_i in the tangent space (Θ_i), projected at the mean of all training trials (black star). To obtain this visualization, the first two principal components of a PCA applied on $\{\Theta_i\}_{i=1}^I$ are selected. In Figures 3b and 3d, a Riemannian potato [46] is applied with the Riemannian mean of all training matrices as a fixed reference. Riemannian potato only removes highly corrupted signals, allowing to remove outlier trials that negatively impact the estimation of class centers. Applying this filter in an offline evaluation improves estimation of class centers for the subject with the lowest BCI performance. Figure 3b illustrates how center estimates are affected by the filtering. Bad estimates of a class center will result in poor clustering.



(a)



(b)

Figure 2: (a) Covariance matrices condition expressed as the ratio \mathcal{C} between largest and smallest eigenvalues for the different covariance estimators. The comparison is done for increasing EEG trial length. (b) Integrated discrimination improvement brought to the classification task by various estimators along varying trail length. The indicated IDI values are multiplied by 10^2 . $\hat{\Sigma}_{scm}$ is used as baseline.

It should be noted that the Riemannian potato may have a negative effect if the different classes are well separated. The Riemannian potato could reject trials that are close to the average of their class but that are

far from the mean of all trials. Thus, “legitimate” trials could be discarded, even if they have a positive impact on the classification results. Avoiding this central bias is a possible improvement of the Riemannian potato algorithm.

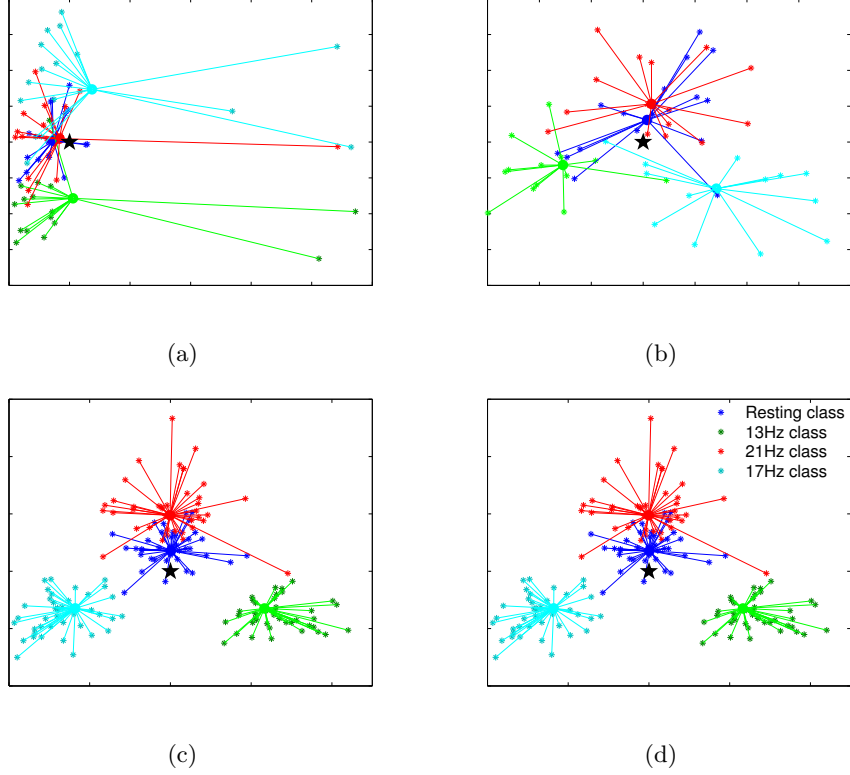


Figure 3: Scatter plot of covariance matrices for all trials mapped on the tangent space. The distance between each trial covariance matrix Σ_i and its Riemannian mean class $\Sigma_\mu^{(k)}$ is shown as connection line. The black star represents the Riemannian mean of all trials. Subject with lowest BCI performance, (a) before and (b) after Riemannian potato filtering. Subject with highest BCI performance, (c) before and (d) after Riemannian potato filtering.

5.4 Classification Results

In this section the classification performances are presented. It is divided into two parts: first, the relevance of identifying the latency between cue onset and SSVEP response is shown through the evaluation of the impact of SSVEP delays considered for offline classification. Then, performances of offline and online algorithms are compared and analysed.

5.4.1 SSVEP Response Time Delay for Offline

Figure 4 shows time signals of one trial from each class, measured at Oz , for the subject with highest BCI performance and for the subject with lowest BCI performance. The figure displays the raw signal filtered at the 3 different stimulus frequency, each with a different color: 13Hz in blue, 17Hz in red, 21Hz in green. It is visible that synchronization for current trial frequency appears only 2 seconds after cue onset $\tau_0 = 0s$ on the x -axis of the plots. In fact, before $\tau_0 + 2$ seconds the signal is still synchronized with the previous trial frequencies. In the signals from the subject with the highest BCI performance, synchronization at the trial stimulus frequency is predominant and permanent after $\tau_0 + 2$, which eases the classification task. Moreover, for the resting trial (no SSVEP), the power in all stimuli frequencies is sensibly decreased. However, the poor predictive capability for the subject with lowest BCI performance could be explained by the fact that the predominance of the trial stimulus frequency is not always evident and not permanent. Also, during the resting trial, the power in stimuli frequencies is not significantly decreased.

Similar observation is made from Figure 5 where the time-frequency representations of EEG trials per class are plotted for the subjects with the highest and lowest BCI performances. The time-frequency representation is obtained with a STFT on rectangular windowing. Each plot has 4 rows, one for each class, displaying the power spectral density (averaged over all trials) in the 13Hz, 17Hz, and 21Hz frequency bands as a function of the trial time. An important increase in average classification accuracy (almost 10%) could be obtained by taking the trial from 2 seconds after cue onset. It is therefore crucial to consider the latency between the trial's cue onset and the actual synchronization of SSVEP at stimulus frequency.

5.4.2 Online Classification Results

In the offline synchronous analysis, the confident window for classification was found to be at least 2 seconds after the cue onset ($\tau_0 + 2$). In an online asynchronous experiment, there is no cue onset, and the delay before SSVEP synchronization might defer from a trial to another and from a subject to another. To locate the trust EEG region for the classification, D and ϑ are set respectively to 5 and 0.7 through cross validation. The epoch size is $w = 3.6$ seconds, and the overlap is $\Delta n = 0.2$ second.

The observation of Figure 6 provides a visualization of the principle guiding the online implementation of Equation (12). This figure shows the trajectory on the tangent space taken by covariance matrices during a 4-class SSVEP experiment, and how they are classified epoch by epoch. It can be seen (encircled in Figure 6a) that a change in the SSVEP stimulus might not be detected instantaneously by the classifier. The trials are er-

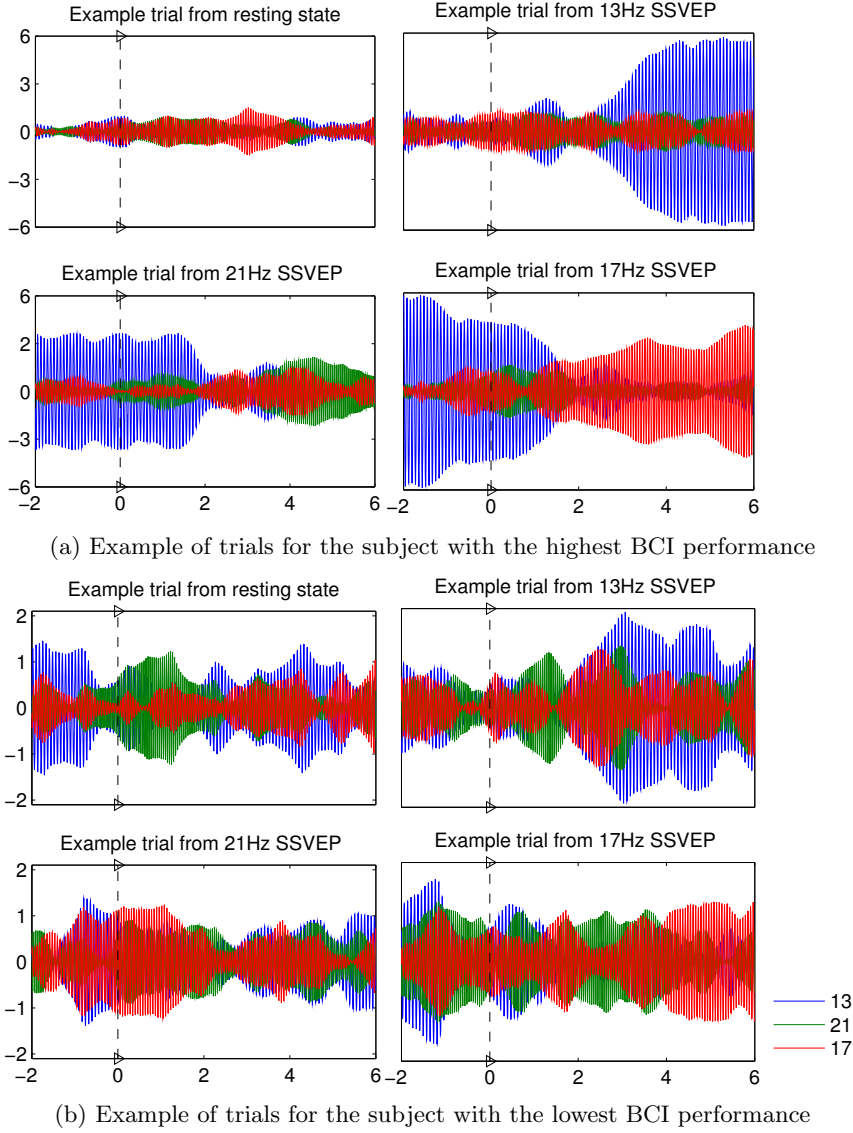
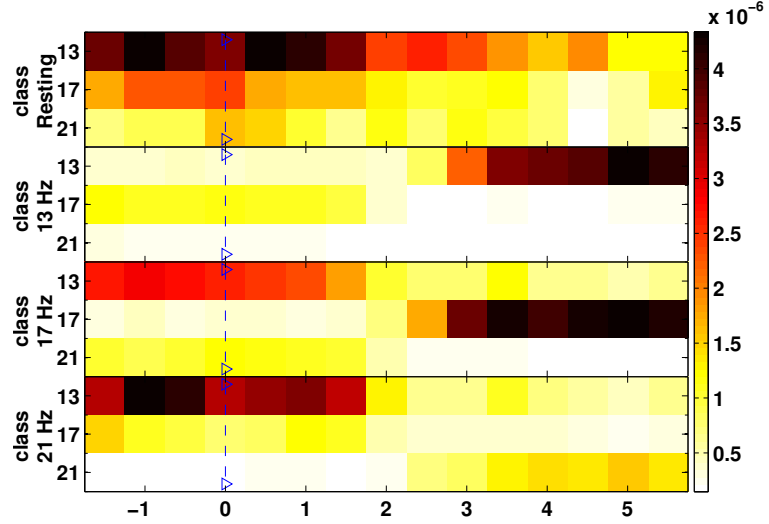
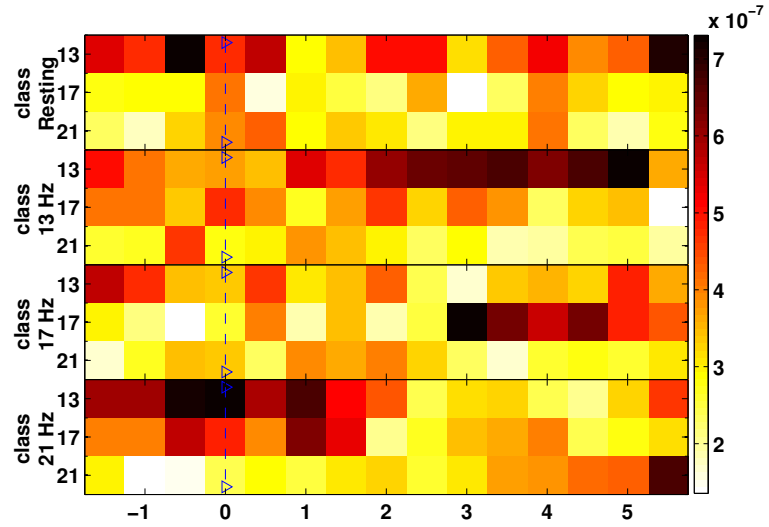


Figure 4: Signal amplitude at each stimulus frequency, showing synchronization of EEG with respect to time (seconds). The raw signal of the trial measured on Oz is band filtered using a Butterworth of order 8 at each stimulus frequency and the resulting signals are shown in blue (dark grey) , green (grey), and red (light grey) for the same signal filtered respectively at 13, 17, and 21 Hz. The cue onset τ_0 at time 0 on the x-axis is shown with a vertical discontinued line. 4 trials are shown, one for each class. Signals from the subjects, (a) with highest BCI performance and (b) with lowest BCI performance.

roneously attributed with confidence to the previous class. The proposed online algorithm, described in Algorithm 3, mitigates this issue and increases the classification accuracy as shown in Table 1. In the ‘Offline’ column are the classification accuracies for each subject obtained by the application of



(a) Time-frequency diagrams for the subject with the highest BCI performance



(b) Time-frequency diagrams for the subject with the lowest BCI performance

Figure 5: Time-frequency representation of EEG from various classes showing synchronization at in frequency bands (indicated on y axis) on both side of the cue onset. The power of synchronization is shown in the color bar. Dark colors show strong synchronization. 4 trials (top to bottom) are represented, one for each class. Signals from the subjects, (a) with highest BCI performance and (b) with lowest BCI performance.

the state-the-art MDRM as described in Algorithm 2, which classifies full

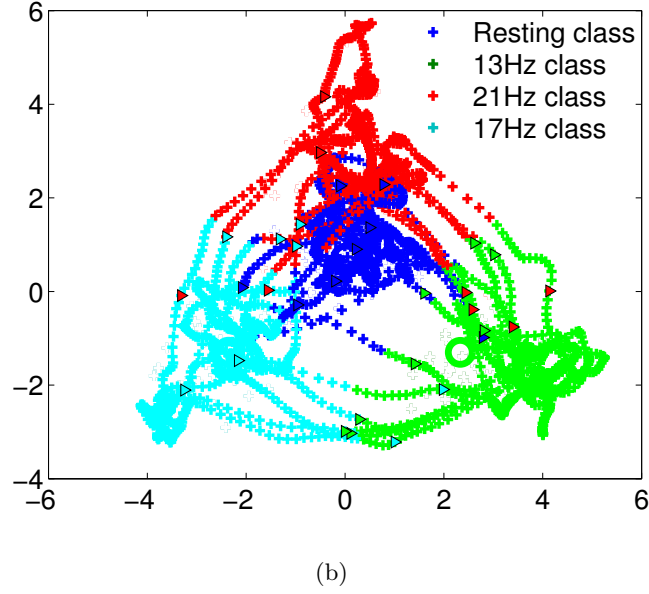
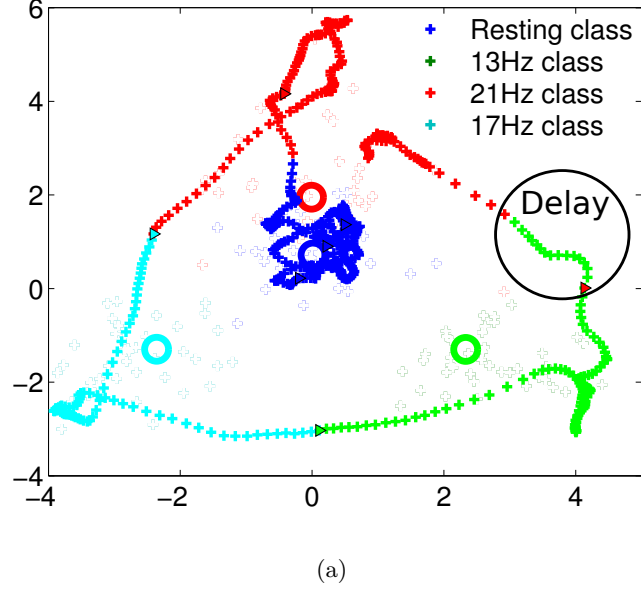


Figure 6: Covariance matrices trajectory during a 4-class SSVEP online recording. The circles represent class centers. The triangles mark the beginning in the experiment of a new trial whose class is indicated by the triangle's color. **6a** shows the first 7 trials. The first 3 trials are from the resting class, the remaining are respectively class 13Hz, 17Hz, and 21Hz. **6b** shows the entire recording. Data are from the subject with the highest BCI performance.

trials. Column 'Offline opt.' shows the results of this approach improved by taking into consideration the SSVEP response time delay as found in

section 5.4.1: trials are taken 2 second after cue onset. Columns ‘Online’ and ‘Online + $\tilde{\delta}_k$ ’ contain the performance of the online algorithm as described in Algorithm 3. The ‘Online’ column shows the results of the online algorithm without the curve direction criterion (*i.e.*, without steps 6 to 11), and ‘Online + $\tilde{\delta}_k$ ’ shows the improvement brought by this criterion. The performances are in terms of average classification accuracy (acc(%)) and average time taken into the trial before classification (delay(s)). For the offline approaches, classification is only done at the end of the trial *i.e.* 6 seconds after cue onset.

The curve direction criterion increases the rejection of epochs that could be wrongly classified, it thus significantly increases the classification accuracy of the online algorithm (74.73% to 79.13%), while increasing the delay (1.006 s to 1.239 s) before classification. When compared to the state-of-the-art MDRM, the curve-based classification yields better results both in terms of classification accuracy and delay before classification, *i.e.* from 71.51% to 79.13% and from 6 s (or longer, depending on trial length N) to 1.239 s. There is a small drop in classification accuracy from optimized offline (81.27%) to online (79.13%) algorithms. However, classification outputs are reached faster with the online algorithm. Moreover, the online algorithm can be applied in both synchronous and asynchronous paradigms, while the offline algorithms are limited to synchronous paradigms which provide strongly limited user interaction.

Sub.	Offline	Offline opt.	Online		Online + $\tilde{\delta}_k$	
	acc(%)	acc(%)	acc(%)	delay(s)	acc(%)	delay(s)
1	62.50	73.44	70.31	1.081	71.37	1.338
2	67.19	79.69	78.12	0.950	75.00	1.303
3	85.94	85.94	87.50	0.950	89.06	0.972
4	78.12	87.50	73.44	1.016	75.00	1.056
5	70.31	68.75	70.31	1.013	70.31	1.319
6	79.69	85.94	87.50	0.997	87.35	1.338
7	78.12	88.54	84.38	1.004	85.42	1.144
8	89.06	92.19	85.94	0.947	89.06	1.072
9	62.50	70.31	67.19	0.922	75.00	1.244
10	60.94	80.47	63.28	1.044	69.53	1.264
11	43.75	65.62	59.38	1.109	68.75	1.375
12	80.00	96.88	69.37	1.036	93.75	1.443
mean	71.51	81.27	74.73	1.006	79.13	1.239

Table 1: Average classification accuracies (acc(%)) achieved using the state-of-the-art MDRM (Offline), the optimized version of the MDRM (Offline opt.), the online algorithm without the curve direction criterion (Online), and the complete online algorithm (Online + $\tilde{\delta}_k$). For the online algorithms, the delay between trial start and classification is also provided (delay(s)).

6 Conclusion

This work investigated the efficiency of Riemannian geometry when dealing with covariance matrices as classification features. A novel algorithm, relying on the curve direction in the space of covariance EEG signals, was introduced and applied on a SSVEP classification task for a 4-class brain computer interface. Different covariance matrix estimators were investigated and their robustness was assessed on multichannel SSVEP signals to ensure that the obtained matrices are accurate estimates of data covariance, are well conditioned, and verify the positive-definiteness property. The Schäfer shrinkage estimator was found to be the best as it yielded the highest classification accuracy with the MDRM algorithm.

The introduced online algorithm enhances the stability of the BCI system, balancing between the classification speed and the prediction accuracy. Classification confidence over a number of epochs mitigates the short term perturbations in the experimental conditions and the attentional variations of the subject, while considering the curve direction overcomes the misclassification of EEG trials that are still synchronized with past stimuli frequencies at classification time.

The Riemannian geometric approach and the determination of the Riemannian mean of classes, offers a powerful tool to develop robust algorithms. A possible extension to the proposed algorithm is to include a dynamic adaptation of the Riemannian mean for each class. This dynamic adaptation was implemented, using the latest successfully classified covariance matrices to replace to oldest ones in the computation of the mean. It is not reported in this paper as there was no significant improvement in the results, which could be attributed to the relatively short duration of the EEG recording sessions, which were no longer than 25 min. The Riemannian means of classes do not change much over the sessions, but for longer recording time this adaptation could bring significant improvements.

Acknowledgment

The authors would like to thank Louis Mayaud from Mensia Technologies for his contribution in discussions that led to the completion of this work.

References

- [1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, T. M. Vaughan, Brain-computer interfaces for communication and control, *Clinical Neurophysiology* 113 (6) (2002) 767–791.

- [2] G. Pfurtscheller, B. Z. Allison, C. Brunner, G. Bauernfeind, T. Solis-Escalante, R. Scherer, T. O. Zander, G. Mueller-Putz, C. Neuper, N. Birbaumer, The hybrid BCI, *Frontiers in neuroscience* 4.
- [3] T. O. Zander, C. Kothe, Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general, *Journal of Neural Engineering* 8 (2) (2011) 025005+.
- [4] J. J. Vidal, Toward direct brain-computer communication., *Annual review of biophysics and bioengineering* 2 (1) (1973) 157–180.
- [5] J. D. Bayliss, D. H. Ballard, Single trial p3 epoch recognition in a virtual environment, *Neurocomputing* 32-33 (2000) 637–642.
- [6] M. A. Lopez, H. Pomares, F. Pelayo, J. Urquiza, J. Perez, Evidences of cognitive effects over auditory steady-state responses by means of artificial neural networks and its use in brain-computer interfaces, *Neurocomputing* 72 (16-18) (2009) 3617–3623.
- [7] Y. Huang, D. Erdogmus, M. Pavel, S. Mathan, K. E. Hild, A framework for rapid visual image search using single-trial brain evoked responses, *Neurocomputing* 74 (12-13) (2011) 2041–2051.
- [8] N. Jrad, M. Congedo, R. Phlypo, S. Rousseau, R. Flamary, F. Yger, A. Rakotomamonjy, sw-SVM: sensor weighting support vector machines for EEG-based brain-computer interfaces, *Journal of Neural Engineering* 8 (2011) 056004+.
- [9] W. Tu, S. Sun, A subject transfer framework for EEG classification, *Neurocomputing* 82 (2012) 109–116.
- [10] E. Niedermeyer, F. Lopes da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 5th Edition, Lippincott Williams & Wilkins, 2004.
- [11] B. Blankertz, K.-R. Müller, D. Krusienski, G. Schalk, J. Wolpaw, A. Schlögl, G. Pfurtscheller, J. Millan, M. Schröder, N. Birbaumer, The BCI competition III: validating alternative approaches to actual BCI problems, *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 14 (2) (2006) 153–159.
- [12] B. Blankertz, K.-R. Müller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schröder, N. Birbaumer, The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials, *Biomedical Engineering, IEEE Transactions on* 51 (6) (2004) 1044–1051.

- [13] B. Blankertz, K. R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. Millan, M. Schroder, N. Birbaumer, The BCI competition III: validating alternative approaches to actual BCI problems, *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 14 (2) (2006) 153–159.
- [14] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. Mueller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, B. Blankertz, Review of the BCI Competition IV, *Frontiers in Neuroscience* 6 (55).
- [15] B. Z. Allison, C. Neuper, Could anyone use a BCI?, in: D. S. Tan, A. Nijholt (Eds.), *Brain-Computer Interfaces, Human-Computer Interaction Series*, Springer London, 2010, Ch. 3, pp. 35–54.
- [16] C. Vidaurre, B. Blankertz, Towards a cure for BCI illiteracy., *Brain topography* 23 (2) (2010) 194–198.
- [17] E. M. Hammer, S. Halder, B. Blankertz, C. Sannelli, T. Dickhaus, S. Kleih, K.-R. Müller, A. Kübler, Psychological predictors of SMR-BCI performance, *Biological Psychology* 89 (1) (2012) 80–86.
- [18] B. Obermaier, C. Guger, C. Neuper, G. Pfurtscheller, Hidden Markov models for online classification of single trial EEG data, *Pattern Recognition Letters* 22 (12) (2001) 1299–1309.
- [19] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, K.-R. Müller, Towards adaptive classification for BCI, *Journal of Neural Engineering* 3 (1) (2006) R13+.
- [20] A. Lenhardt, M. Kaper, H. Ritter, An Adaptive P300-Based Online Brain Computer Interface, *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 16 (2) (2008) 121–130.
- [21] G. Bin, X. Gao, Z. Yan, B. Hong, S. Gao, An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method, *Journal of Neural Engineering* 6 (4).
- [22] D. Zhu, G. Garcia-Molina, V. Mihajlović, R. Aarts, Online BCI implementation of High-Frequency phase modulated visual stimuli, in: C. Stephanidis (Ed.), *Universal Access in Human-Computer Interaction. Users Diversity*, Vol. 6766 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2011, pp. 645–654.
- [23] H. Lee, S. Choi, PCA+HMM+SVM for EEG pattern classification, *Signal Processing and Its Applications*, 2003. *Proceedings. Seventh International Symposium on* 1 (2003) 541–544.

- [24] E. Kalunga, K. Djouani, Y. Hamam, S. Chevallier, E. Monacelli, SSVEP enhancement based on Canonical Correlation Analysis to improve BCI performances, in: AFRICON, 2013, IEEE, 2013, pp. 1–5.
- [25] H. Lu, H.-L. Eng, C. Guan, K. Plataniotis, A. Venetsanopoulos, Regularized Common Spatial Pattern With Aggregation for EEG Classification in Small-Sample Setting, *Biomedical Engineering, IEEE Transactions on* 57 (12) (2010) 2936–2946.
- [26] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K. R. Muller, Optimizing Spatial filters for Robust EEG Single-Trial Analysis, *Signal Processing Magazine, IEEE* 25 (1) (2008) 41–56.
- [27] F. Lotte, C. Guan, Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms, *Biomedical Engineering, IEEE Transactions on* 58 (2) (2011) 355–362.
- [28] K. K. K. Ang, Z. Y. Y. Chin, C. Wang, C. Guan, H. Zhang, Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b., *Frontiers in neuroscience* 6.
- [29] Y. Yang, S. Chevallier, J. Wiart, I. Bloch, Automatic selection of the number of spatial filters for motor-imagery BCI, in: M. Verleysen (Ed.), *European Symposium on Artificial Neural Networks (ESANN)*, 2012, pp. 109–114.
- [30] D. R. Hardoon, S. R. Szedmak, J. R. Shawe-Taylor, Canonical Correlation Analysis: An Overview with Application to Learning Methods, *Neural Comput.* 16 (12) (2004) 2639–2664.
- [31] Z. Lin, C. Zhang, W. Wu, X. Gao, Frequency Recognition Based on Canonical Correlation Analysis for SSVEP-Based BCIs, *Biomedical Engineering, IEEE Transactions on* 53 (12) (2006) 2610–2614.
- [32] P.-A. Absil, R. Mahony, R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2009.
- [33] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Riemannian geometry applied to BCI classification, in: *Latent Variable Analysis and Signal Separation*, Springer, 2010, pp. 629–636.
- [34] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Multiclass brain–computer interface classification by Riemannian geometry, *Biomedical Engineering, IEEE Transactions on* 59 (4) (2012) 920–928.
- [35] R. C. Panicker, S. Puthusserypady, Y. Sun, Adaptation in p300 brain–computer interfaces: A two-classifier cotraining approach, *Biomedical Engineering, IEEE Transactions on* 57 (12) (2010) 2927–2935.

- [36] F. Schettini, F. Aloise, P. Aricò, S. Salinari, D. Mattia, F. Cincotti, Self-calibration algorithm in an asynchronous p300-based brain-computer interface, *Journal of neural engineering* 11 (3) (2014) 035004.
- [37] H. Verschore, P.-J. Kindermans, D. Verstraeten, B. Schrauwen, Dynamic stopping improves the speed and accuracy of a P300 speller, in: *Artificial Neural Networks and Machine Learning-ICANN 2012*, Springer, 2012, pp. 661–668.
- [38] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, et al., BCI signal classification using a riemannian-based kernel, in: *Proceeding of the 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2012, pp. 97–102.
- [39] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Classification of covariance matrices using a riemannian-based kernel for BCI applications, *Neurocomputing* 112 (2013) 172–178.
- [40] F. Yger, A review of kernels on covariance matrices for BCI applications, in: *Machine Learning for Signal Processing (MLSP)*, 2013 IEEE International Workshop on, IEEE, 2013, pp. 1–6.
- [41] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, M. Harandi, Kernel methods on the riemannian manifold of symmetric positive definite matrices, in: *Computer Vision and Pattern Recognition (CVPR)*, 2013 IEEE Conference on, IEEE, 2013, pp. 73–80.
- [42] Y. Xie, J. Ho, B. Vemuri, On a nonlinear generalization of sparse coding and dictionary learning, in: *Proceedings of the 30th International Conference on Machine Learning*, NIH Public Access, 2013, p. 1480.
- [43] A. Goh, R. Vidal, Unsupervised Riemannian clustering of probability density functions, in: *Machine Learning and Knowledge Discovery in Databases*, Springer, 2008, pp. 377–392.
- [44] A. Goh, R. Vidal, Clustering and dimensionality reduction on Riemannian manifolds, in: *Computer Vision and Pattern Recognition*, 2008. CVPR 2008. IEEE Conference on, IEEE, 2008, pp. 1–7.
- [45] X. Pennec, P. Fillard, N. Ayache, A riemannian framework for tensor computing, *International Journal of Computer Vision* 66 (1) (2006) 41–66.
- [46] A. Barachant, A. Andreev, M. Congedo, et al., The Riemannian potato: an automatic and adaptive artifact detection method for online experiments using Riemannian geometry, in: *Proceedings of TOBI Workshop IV*, 2013, pp. 19–20.

- [47] S.-I. Amari, Information geometry in optimization, machine learning and statistical inference, *Frontiers of Electrical and Electronic Engineering in China* 5 (3) (2010) 241–260.
- [48] W. Samek, D. Blythe, K.-R. Müller, M. Kawanabe, Robust spatial filtering with beta divergence, in: *Advances in Neural Information Processing Systems*, 2013, pp. 1007–1015.
- [49] W. Samek, K.-R. Müller, Information geometry meets bci spatial filtering using divergences, in: *Brain-Computer Interface (BCI), 2014 International Winter Workshop on*, IEEE, 2014, pp. 1–4.
- [50] A. Barachant, S. Bonnet, Channel selection procedure using Riemannian distance for BCI applications, in: *Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on*, IEEE, 2011, pp. 348–351.
- [51] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Common spatial pattern revisited by Riemannian geometry, in: *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, IEEE, 2010, pp. 472–476.
- [52] M. Congedo, A. Barachant, A. Andreev, A new generation of brain-computer interface based on riemannian geometry, *arXiv preprint arXiv:1310.8115*.
- [53] Y. Li, K. M. Wong, H. De Bruin, Eeg signal classification based on a riemannian distance measure, in: *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*, IEEE, 2009, pp. 268–273.
- [54] Y. Li, K. Wong, H. De Bruin, Electroencephalogram signals classification for sleepstate decision-a riemannian geometry approach, *Signal Processing, IET* 6 (4) (2012) 288–299.
- [55] J. Lee, *Introduction to topological manifolds*, Vol. 940, Springer, 2010.
- [56] J. Jost, *Riemannian geometry and geometric analysis*, Vol. 62011, Springer, 2011.
- [57] M. Moakher, A differential geometric approach to the geometric mean of symmetric positive-definite matrices, *SIAM Journal on Matrix Analysis and Applications* 26 (3) (2005) 735–747.
- [58] P. T. Fletcher, C. Lu, S. M. Pizer, S. Joshi, Principal geodesic analysis for the study of nonlinear statistics of shape, *Medical Imaging, IEEE Transactions on* 23 (8) (2004) 995–1005.

- [59] K. Fukunaga, Introduction to statistical pattern recognition, Academic press, 1990.
- [60] O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, *Journal of multivariate analysis* 88 (2) (2004) 365–411.
- [61] B. Blankertz, S. Lemm, M. Treder, S. Haufe, K.-R. Müller, Single-trial analysis and classification of erp componentsa tutorial, *NeuroImage* 56 (2) (2011) 814–825.
- [62] J. Schäfer, K. Strimmer, A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics, *Statistical applications in genetics and molecular biology* 4 (1).
- [63] F. Pascal, P. Forster, J. P. Ovarlez, P. Arzabal, Theoretical analysis of an improved covariance matrix estimator in non-gaussian noise, in: *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*., Vol. 4, 2005.
- [64] M. J. Pencina, R. B. D’Agostino, R. S. Vasan, Evaluating the added predictive ability of a new marker: from area under the roc curve to reclassification and beyond, *Statistics in medicine* 27 (2) (2008) 157–172.